National Aeronautics and Space Administration

Algorithm Theoretical Basis Document (ATBD) for Global Precipitation Climatology Project Version 3.2 Daily Precipitation Data

Prepared by:

George J. Huffman
NASA/GSFC
NASA/GSFC Code 612
Greenbelt, MD 20771

and Robert F. Adler, Ali Behrangi, David T. Bolvin, Eric J. Nelkin, Mohammad Reza Ehsani

Last Revised January 24, 2023

Revision History

Revision Date	*Changes*	*Author*
June 9, 2021	Original (Version 3.1)	George Huffman
February 1, 2022	Version 3.2	George Huffman
January 24, 2023	Extend record to September 2021	George Huffman

Table of Contents

Contents

Tables

Figures

1.0 Introduction

Precipitation observations are critical to many applications including drought monitoring, flash floods, crop forecasting, disease prediction, and ocean salinity studies. Surface precipitation gauges (hereafter simply "gauges") are the primary source of direct precipitation observations. Unfortunately, these gauges are point measurements and much of the globe is sparsely covered, especially in underdeveloped countries and areas of low population density. Furthermore, with the exception of a few buoy arrays, there are no precipitation gauge observations over the open ocean. Satellites mitigate the limitations of rain gauge observations by providing precipitation estimates over land and ocean for most, or all, of the entire globe. When converted to gridded precipitation estimates, the satellite observations facilitate a multitude of studies, including those on the larger space-time scales that gauge analyses typically cannot provide. To augment the satellite-based precipitation estimates, uniformly processed gauge analyses are incorporated to improve the land-based estimates.

The Global Precipitation Climatology Project (GPCP) is a community-based activity supported by the Global Water and Energy Exchange (GEWEX) project of the World Climate Research Programme (WCRP), focused on creating a global, long-term homogeneous record of gridded precipitation estimates and ancillary information for use in climate studies and other applications. GPCP Version 3 is the successor to the highly successful GPCP V2 data set. The current GPCP V3.2 products are at the monthly and daily resolution with the 3-hourly products to be developed next.

1.1 Purpose

The purpose of this document is to describe the algorithm used to create the Global Precipitation Climatology Project (GPCP) Version 3.2 Daily Satellite analysis. It was developed primarily under a current NASA Making Earth Science Data Records for Use in Research Environments (MEaSUREs) program (PI: Ali Behrangi, University of Arizona). Prior versions of the GPCP analysis have been produced by a consortium of individual scientists at various government and university institutions and most recently as part of the NOAA Climate Data Record (CDR) Program. The current GPCP Daily product described here creates a global daily satellite product using the Integrated Multi-satellitE Retrievals for the Global Precipitation Measurement (GPM) mission (IMERG) half-hourly Final Run product (accumulated to daily) at low and middle latitudes, and daily retrievals from Television InfraRed Operational Satellite (TIROS) Operational Vertical Sounder (TOVS) and Atmospheric Infrared Sounder (AIRS) at high latitudes, with calibration to the GPCP V3.2 Monthly SG product. The intent here is to provide a guide to understanding the algorithm from a scientific perspective.

1.2 Definitions

Symbols and acronyms used in the document are defined and summarized in Appendix A.

1.3 Document Maintenance

This document describes the Version 3.2 GPCP Daily satellite product.

1.4 What's New?

V3.2 Daily is the second release in Version 3 of daily data, intended to eventually replace the previous V1.3 One-Degree Daily (1DD). The team continues to work toward improving the processing for future releases and extend back in time to match the 1983 start for V3.2 Monthly (Huffman et al. 2022).

2.0 Observing Systems Overview

2.1 Products Generated

This document describes the GPCP Daily satellite dataset. The primary output of this algorithm is daily precipitation starting in June 2000 on a 0.5°, globally complete grid obtained by calibrating daily precipitation observations from satellites to the monthly SG product. The prior GPCP daily analysis procedures are described in Huffman et al. (2001). In addition to the daily precipitation analysis, the product provides precipitation phase.

2.2 Instrument Characteristics

The GPCP Daily satellite precipitation product is based on data from polar orbiting satellites (directly and through IMERG), geostationary satellites (through IMERG), and gauges (through IMERG and the Monthly SG calibration). The actual data used are described in section 3.3.1. The following gives information on satellite sensor characteristics that are used directly.

TOVS

The Television InfraRed Operational Satellite (TIROS) Operational Vertical Sounder (TOVS) dataset of surface and atmospheric parameters is derived from analysis of High-Resolution Infrared Sounder 2 (HIRS2) and Microwave Sounding Unit (MSU) data aboard the NOAA series of polar-orbiting operational meteorological satellites. The precipitation estimates from TOVS are derived as a secondary product utilizing various retrieved sounding parameters, including atmospheric temperature and water vapor profiles, cloud-top pressure, and radiatively effective fractional cloud cover.

For the period January 1979 - February 1999 (not currently used), the TOVS estimates are based on two NOAA satellites. Beginning in March 1999 (used June 2000 – August 2002), the TOVS estimates are based on a single NOAA satellite due to the failure of NOAA-11. In addition, the available TOVS record has a processing change starting with December 1999 (uncovered in developing V3.1) as described in Section 3.3.1 in "Adjustments to TOVS precipitation". More information can be found in Susskind et al. (1997).

AIRS

The Atmospheric Infrared Sounder (AIRS) aboard the NASA Aqua polar-orbiting satellite is the source of precipitation estimates that succeeded TOVS (in September 2002, although TOVS continued to produce successively more-degraded observations into 2005). The precipitation estimates from AIRS are derived in a very similar way to those from TOVS as a secondary product utilizing various retrieved sounding parameters, including atmospheric temperature and water vapor profiles, cloud-top pressure and radiatively effective fractional cloud cover (Susskind et al. 2003). Because the Advanced Microwave Sounding Unit (AMSU) aboard Aqua failed in late September 2016, GPCP Version 3.2 Daily uses AIRS precipitation values for the entire record that were reprocessed and are based on infrared-only data for the sake of homogeneity.

3.0 Algorithm Description

3.1 Algorithm Overview

The algorithm to produce the Daily 0.5° GPCP product takes inputs from several different sources and merges them to create the most consistent and accurate daily precipitation estimates. This document describes the high-level procedures, inputs, and outputs, of the GPCP V3.2 Daily analysis. Many parts of the GPCP V3.2 Daily techniques are built upon concepts used in GPCP V1 1DD, the details of which can be found in Huffman et al. (2001). IMERG, TOVS, and AIRS are satellite-based precipitation estimates computed outside the merger process described here. The GPCP V3.2 Monthly product is input as well. The IMERG data are used in low and middle latitudes, with TOVS/AIRS filling in the high latitudes, then each grid box's time series for the month is scaled to approximately sum to the monthly value.

3.2 Processing Outline

Figure 1 shows the steps required for GPCP Daily processing. Over the period of the GPCP V3.2 Daily (June 2000 to near-present) there was a change from TOVS to AIRS in September 2002, but calibrations between the records were made outside this processing, so no additional work is required here.

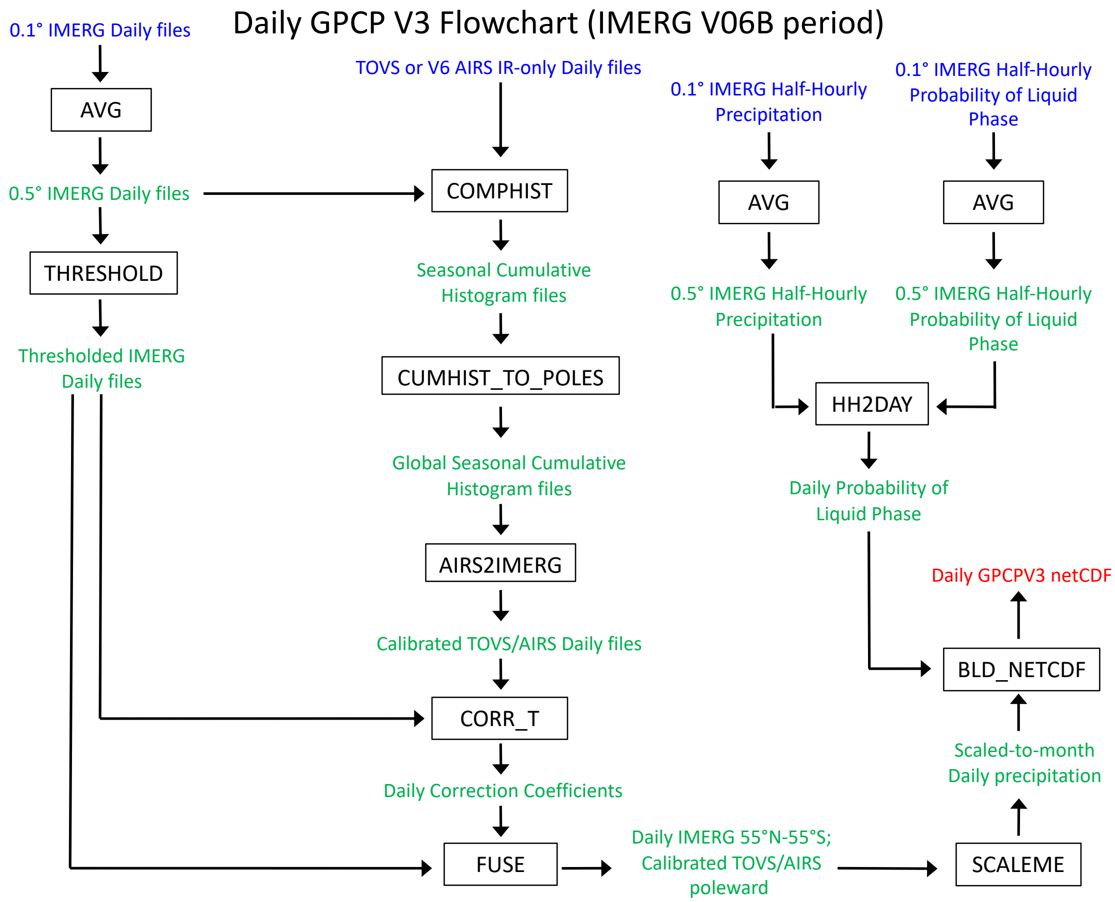

Figure 1: Processing steps required to produce GPCP Daily. Names in boxes approximate names used by code.

3.3 Algorithm Input

3.3.1 Primary Sensor Data

No primary sensor data are used in the V3.2 Daily. The IMERG Final, TOVS, and AIRS daily estimates and GPCP V3.2 Monthly estimates all enter the analysis as pre-computed precipitation fields (Fig. 2).

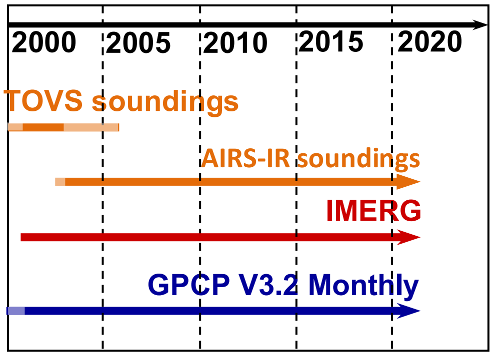

Figure 2: Inputs to GPCP V3.2 Daily by time. Lighter line segments show data periods not used in V3.2 processing for the various inputs.

Table 1: Inputs to GPCP Daily.

Name	Data type	GPCP time period	Data source	Notes
GPCP V3.2 Monthly	Precip	June 2000 to present	GSFC/GPCP	Companion Monthly product
TOVS	Precip	June 2000 to Aug 2002	GSFC/SRT	
AIRS IR-only V6	Precip	Sep 2002 to present	GES DISC	Replaces TOVS
IMERG V6 Final half-hourly	Precip	June 2000 to present	GES DISC	

TOVS precipitation

The TOVS instrument flew aboard the NOAA series of polar-orbiting platforms. Susskind and Pfaendtner (1989) and Susskind et al. (1997) described the process for estimating precipitation from TOVS. The TOVS precipitation estimates infer precipitation from deep, extensive clouds. The technique begins with a first guess driven by a simple global numerical analysis, then uses a climatological multiple regression relationship between collocated First Global Atmospheric Research Program (GARP) Global Experiment (FGGE) precipitation gauge measurements and several TOVS-based parameters that relate to cloud volume: cloud-top pressure, fractional cloud cover, and relative humidity profile. This relationship is allowed to vary seasonally and latitudinally. Furthermore, separate relationships are developed for ocean and land.

The TOVS data are used June 2000-August 2002 and are provided as daily 1° gridded estimates for October 1996 – August 2002. The data are based on information from one satellite due to changes in satellite data format. TOVS data were obtained directly from the NASA Goddard Sensor Research Team (SRT), led by Joel Susskind.

AIRS precipitation

The AIRS instrument is flying aboard the Earth Observing System Aqua polar-orbiting satellite and has been used to succeed TOVS data. The same algorithm applied to TOVS to produce precipitation is also applied to AIRS (Susskind et al. 2003), except the first guess is a spatially and seasonally varying climatology. The AIRS data are provided in daily 0.25° gridded format, then averaged to 0.5° resolution. AIRS data are available from September 2002 and used for the period from September 2002 – present. [The TOVS data continued to May 2005, but are considered questionable after early 2003.]

Adjustments to TOVS precipitation

To adjust the TOVS estimates to AIRS to maintain a homogeneous record, the TOVS data were first "zoomed" from 1° resolution to 0.5° via grid box replication. A histogram-matching approach was then used to calibrate the TOVS estimates to AIRS. Histograms were developed using a 24-month period of daily data. We chose to create the calibration using independent data due to the very limited overlap period between TOVS and AIRS (September-December 2002). Priority for the 24 months was assigned to selecting two of each calendar month, then to the more ENSO-neutral conditions among the available choices. Since the resulting histogram relationships are applied to the TOVS 1-satellite period (that began in March 1999), the 24 months of TOVS used to build the histograms were chosen from the 1-satellite period. These months are July 2000-June 2002. The 24 AIRS months of daily data consisted of September 2012-August 2014, except March 2015 replaced March 2014 due to six missing days in the latter. Note that the shift in processing of the TOVS input detected in developing V3.1 occurs in December 1999, before the start of the V3.2 record.

One global land histogram was generated for those grid boxes containing 0-15% ocean. One global "mixed" histogram was generated for grid boxes containing 15-75% ocean. For grid boxes with 75+% ocean, a set of 34 histograms were generated, in 15° latitude bands with 10-degree overlap (i.e., 90°-75°N, 85°-70°N, …, 75°-90°S). The resulting adjustment was then applied to the period December 1999-August 2002, and these daily adjusted TOVS files were accumulated to monthly.

As the final step, the entire daily TOVS/AIRS record is scaled by the monthly climatological Merged CloudSat, Tropical Rainfall Measurement Mission (TRMM), and GPM Climatology (MCTG; see Behrangi and Song 2020, Behrangi et al. 2012, 2014) ratios, capped at a maximum value of 2, to obtain the best estimate of precipitation using TRMM, GPM, and CloudSat information. This is especially important at latitudes higher than 55°N-S where TOVS/AIRS is the only source of satellite precipitation estimates in V3.2 Daily and the CloudSat calibration dominates. The MCTG is described later in this section.

IMERG precipitation

IMERG assembles the time-varying constellation of available passive microwave (PMW) satellite sensors, intercalibrated to the GPM Microwave Imager (GMI) as GPM Level 1C brightness temperatures (Tc), then converted to Level 2 precipitation estimates using the V05 Goddard Profiling (GPROF; Kummerow et al. 2015) scheme. All estimates are gridded, inter-calibrated to the Ku swath Combined Radar-Radiometer (CORRA; Olson et al. 2011) product on a rolling 45-day basis using probability matching, and climatologically calibrated to the GPCP V2.3 Monthly SG estimates with a simple ratio in high latitude oceans (where CORRA is deficient in precipitation) and over all land areas (where CORRA tends to be high). In the TRMM era, TMI and CORRA-TRMM (CORRA-T) are used. These PMW-based precipitation data are combined into half-hourly fields, masked for surface snow and ice (due to uncertain quality in GPROF), and provided to both the recalibration of Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks – Cloud Classification System (PERSIANN-CCS; Hong et al. 2004) infrared estimates and to the semi-Lagrangian time interpolation scheme adapted from Climate Prediction Center (CPC) Morphing-Kalman Filter (CMORPH-KF; Joyce et al. 2011). Modern Era Retrospective Reanalysis 2 (MERRA-2; Gelaro et al. 2017) numerical analysis fields of vertically integrated water vapor are used in the CMORPH-KF semi-Lagrangian time interpolation scheme, and the CPC assembles the zenith-angle-corrected, inter-calibrated "even-odd" and merged geo-IR fields for use in the PERSIANN-CCS computations. The PERSIANN-CCS estimates are computed (supported by an asynchronous 30-day re-calibration cycle) and sent to the CMORPH-KF weighting scheme. The CMORPH-KF weighting scheme (supported by an asynchronous KF weights 3-monthly updating cycle) uses the PMW and IR estimates to create half-hourly estimates. The IMERG product used here is the Final satellite-gauge product, in which the half-hour multi-satellite estimates are adjusted to equal the monthly satellite-gauge combination computed in a separate monthly IMERG estimate. Note that each IMERG version, including V06, is uniformly processed over the entire record (currently June 2000 – September 2021), although changes in the available satellites lead to changes in dataset quality and performance. See Huffman et al. (2020) for more information.

MCTG Climatology

The MCTG is a composite monthly climatology based on CloudSat, TRMM-and GPM-based precipitation estimates using a concept similar to that described in Behrangi and Song (2020) and Behrangi et al. (2012; 2014). CloudSat precipitation frequency is first adjusted for the spatial resolution of TRMM and GPM radar footprint (~5km) and is used as a constraint for combining CloudSat precipitation (rainfall plus snowfall), precipitation rates from the TRMM Combined Climatology (TCC; within 25°N-S), the TRMM Level 2 CORRA-T (product 2BCMBT V06; within 35°N-S), and the GPM Level 2 CORRA (product 2BCMB V06; within 35° – 65°N-S). Poleward of 65°N-S the precipitation climatology is based on CloudSat as it does not have a signal saturation issue (Behrangi et al. 2012) and GPM lacks coverage. The GPM 2BCMB is used as it provides an overlapping period with CloudSat (2007-2010) when both day- and night-time observations are available.

Note that the TOVS/AIRS estimates are globally adjusted to the MCTG prior to use in V3.2 Daily to take advantage of CloudSat climatological information at high latitudes.

GPCP V3.2 precipitation

Briefly, six relatively homogeneous data sets form the basis of the GPCPV3.2 Monthly product. Passive microwave (PMW) radiometer estimates are computed from the Defense Meteorological Satellite Program (DMSP) Special Sensor Microwave/Imager (SSMI) that flew on the F11 and F13 spacecraft, and Special Sensor Microwave Imager/Sounder (SSMIS) that is flying on the F17 spacecraft, all using the GPROF 2010 Version 2 (GPROF2010v2; Kummerow et al. 2011) and the Microwave Emission Brightness Temperature Histograms (METH; Chiu et al. 1993, Chiu and Chokngamwong 2010) computed as part of GPCP V2.3. Infrared data from the global collection of geosynchronous weather satellites are converted to precipitation estimates using the PERSIANN Climate Data Record (PERSIANN-CDR; Ashouri et al. 2015). The NASA/GSFC Sounder Research Team provides precipitation estimates from the TOVS sensors that flew on selected TIROS- and NOAA-series satellites, and on the AIRS instrument aboard the Earth Observing System Aqua satellite (Susskind and Pfaendtner 1989; Susskind et al. 1997; Susskind et al. 2003; Susskind et al. 2014). The TCC is provided by the University of Maryland, and the MCTG climatology is provided by the University of Arizona. Finally, global gauge analyses are provided by the Deutscher Wetterdienst (DWD) Global Precipitation Climatology Centre (GPCC; Schneider et al. 2017; Schneider et al. 2014; Becker et al. 2013).

These data are merged via a set of algorithms to take advantage of the strengths of each data set and minimize the weaknesses to create a single, best precipitation estimate with associated ancillary fields. Sparse, high-quality GPROF precipitation estimates calibrate the more frequent, near-global (60°N–60°S) PERSIANN-CDR estimates, which are then adjusted by monthly climatological blended TCC/MCTG ratios. Starting at 58° N and S the TOVS/AIRS estimates, adjusted by monthly climatological MCTG ratios, are used at the high latitudes (58°–90° N and S). The Legates and Wilmott (1990) wind-loss correction is applied to the GPCC gauge analysis, with a climatology of the Fuchs et al. (2001) correction used over Eurasia poleward of 45°N. Because the GPROF estimates are currently only continuously available starting in January 1992, the period January 1983 – December 1991 is based on PERSIANN-CDR that has been seasonally climatologically calibrated to the PMW-calibrated PERSIANN-CDR for the overlap period January 1993 – December 2018. Next, the global satellite-only estimates and the global wind-loss-corrected GPCC gauge analysis are merged as 1) adjust the satellite-only estimate to the large-scale mean of the gauge analysis, then 2) merge the adjusted satellite-only estimate with the GPCC gauge analysis based on inverse error variance to produce the final precipitation estimate. Details of the algorithm can be found in Huffman et al. (2022).

3.4 Theoretical Description

The bulk of the GPCP code merges the various inputs to obtain the GPCP Daily estimate. Most of the input data are obtained as rain rates having been processed by other groups. The exception to this is the IMERG Probability of Liquid Phase (PLP), which is a probability.

3.4.1 Physical and Mathematical Description

No precipitation estimates are computed as part of the GPCP V3.2 Daily.

3.4.2 Data Merging Strategy

The V3.2 GPCP Daily is created by combining and calibrating precipitation estimates.

Daily IMERG

The original half-hourly 0.1°x0.1° IMERG Final Run data are accessed as daily and averaged to 0.5°x0.5°.

IMERG-calibrated TOVS and AIRS

The TOVS record was calibrated to the AIRS record off-line, so the two sources are considered to provide an approximately uniform record. TOVS/AIRS is adjusted to IMERG using regionally (3x3 template of the 1°x1° gridboxes) and seasonally varying histograms of daily precipitation rates. In the polar regions where IMERG lacks data (either seasonally or persistently), the matched histograms are smooth-filled[1] to provide calibration for TOVS/AIRS. As well, the IMERG-calibrated TOVS/AIRS is bias-adjusted to the GPCP V3.2 Monthly to ensure a reasonable bias for joining to the IMERG data region (below).

Combined Satellite Precipitation

Preliminary work showed that the fine temporal sampling in IMERG tended to depict too much light precipitation, resulting in a mis-match with the TOVS/AIRS data, so the daily IMERG is thresholded at 0.24 mm/d. Then IMERG is used in the latitude band 55°N-S, and TOVS/AIRS is used at higher latitudes.

The boundary shows some discontinuities in some daily precipitation fields, so a light, empirically developed “feathering” is applied. The 3x3 gridbox smoothed difference (IMERG – TOVS/AIRS) is computed just inside the 55° N and S boundaries, then extended into the TOVS/AIRS region. For the extension, as each higher-latitude gridbox is encountered, the centered average of the 3-gridbox difference values at the previous latitude is computed, forced toward zero by 2 mm/day (determined in a preliminary analysis). [This approach forces small differences to zero in a smaller interval than larger differences.] If the difference is still non-zero after 5° of latitude, the differences over the entire 5° are raised to force a zero difference after 5°. The balance in this scheme is between discontinuities in the daily fields at 55° N and S and changes (usually increases) in the fractional coverage by precipitation in the feathering region.

---

[1] Smooth-filling is an iterative process. On each pass, the value in every gridbox that originally was “missing” is replaced by the average of (non-missing) values on the stated template (here, 3 gridboxes in the X and Y directions). This continues until the data field (approximately) converges.

Finally, for each gridbox, all the days are ratio-adjusted to approximately sum to the GPCP V3.2 Monthly. Since the Monthly is strongly influenced by the precipitation gauge analysis, the adjustment ratios for the Daily are capped in the range 0.2-4 to prevent mismatches between precipitation gauge and satellite values from driving unrealistically high rates for the few days of occurrence in regions where precipitation events are sparse.

3.4.3 Precipitation Phase

The various input precipitation estimates provide total hydrometeor mass in the atmospheric column and then implicitly correlate it to surface precipitation in any phase. Given this fact, the "precipitation" reported in this document refers to all forms of precipitation, including rain, drizzle, snow, graupel, and hail.

Since the precipitation phase, namely whether it is liquid, solid, or mixed, is not currently provided as a satellite-based calculation by the precipitation algorithms used in GPCP V3.2, we must use ancillary data sets to create the estimate. Formally, there should be separate estimates for each phase. However, mixed-phase cases tend to be a small fraction of all cases (except perhaps in the Southern Ocean), and we consider the estimation schemes to be sufficiently simplistic that estimating mixed phase as a separate class seems unnecessary. Some users need information on the occurrence of the solid phase, both due to the delays it introduces in moving precipitation water mass through hydrological systems, and due to the hazardous surface conditions that snow and ice create. Since mixed precipitation frequently melts soon after falling, we lump together liquid and mixed as "liquid" and compute a simple probability of liquid phase. Another simplification is that the IMERG datasets provide globally complete PLP estimates on a 0.1°x0.1° grid every half hour as a diagnostic specification using surface type, surface pressure, surface temperature, and surface humidity (after Sims and Liu 2015) as provided by European Centre for Medium-Range Weather Forecasting (ECMWF) Reanalysis 5 (ERA5; Hersbach et al. 2020). Since this diagnostic only depends on the ancillary data, it is globally complete, providing values even when IMERG does not provide precipitation estimates (specifically in high latitudes).

At the daily scale the probability could either be the fraction of the time that the precipitation is liquid or the fraction of the daily accumulation that fell as liquid. The latter seems to be what most users will want, so this is the parameter computed. The daily PLP is computed as the precipitation-rate-weighted average of all half-hourly probabilities in the day, except for grid boxes where zero precipitation is estimated for the day, in which case it is the simple average of all available probabilities in the day. This approach assumes that the occurrence of liquid and solid over the day will approximately conform to the percentages given in the specification equation, so that the weighted PLP approximates the fraction of amount of precipitation: liquid precipitation = probability * precipitation, and solid precipitation = (100 – probability) * precipitation. In the regions where (daily) TOVS/AIRS estimates are used, the daily PLP is the simple average of the 48 half-hourly PLP in the day since IMERG is not available to provide weighting in parts of that region.

Note that the assignment of phase does not change the units of precipitation, which is the depth of liquid. In the case of solid precipitation, this is usually referred to as snow water equivalent (SWE). The depth of fallen snow that corresponds to this SWE depends on the density of the snow. Typically, it takes about 10 mm of fallen snow to yield 1 mm of SWE, but the ratio depends on location, meteorological regime, time of year, and elevation. There is an excellent discussion of how Environment Canada is addressing this in Wang et al. (2017).

3.4.4 Algorithm Output

The GPCP Version 3.2 Daily satellite precipitation data set covers the period June 2000 throughSeptember 2021). The primary product in the dataset is a combined observation-only dataset, that is, a gridded analysis based on satellite estimates of precipitation and (indirectly) gauge measurements.

The data set archive consists of daily netCDFs, with each file having the following two fields:

(1) merged satellite-gauge precipitation estimate (mm/d),
(2) probability of liquid phase (%).

Each file occupies almost 1.6 MB. The grid on which each field of values is presented is a 0.5°x0.5° latitude–longitude (Cylindrical Equal Distance) global array of points. It is size 720x360, with X (longitude) incrementing most rapidly West to East from the International Dateline, and then Y (latitude) incrementing North to South. Grid edges are placed on whole- and half-degree values:

First point center = (89.75°N, -179.75°W)
Second point center = (89.75°N, -179.25°W)
Last point center = (89.75°S, 179.75°E)

4.0 Test Datasets and Output

4.1 Test Input Datasets

Initial test input data sets consisted of several months of data inputs. Major development work was done with January 2018. Completed datasets for July 2017, October 2017, and April 2018 were reviewed for consistency. November 2003 was also reviewed to make sure the prolonged AIRS outage in that month was handled correctly.

4.2 Test Output Analysis

4.2.1 Precision and Accuracy

It is a matter of research to estimate the random error in the Daily product. For the latitude band 55°N-S its random error characteristics are close to those of the IMERG daily, while at higher latitudes it is similar to the previous V1.3 1DD. Qualitatively, the random error is a strong function of precipitation rate, and varies by climate zone. Overall, there is a fair amount of scatter, given the relatively fine spatial scale. Users should consider performing averaging appropriate to their particular application to reduce these random variations. Section 3.4.2 uses the MCTG climatology to approximately adjust the GPCP V3.2 daily estimates to what is considered the “best” available climatology, but this remains a research topic.

Probability of liquid phase (PLP) is in units of whole percents. Tests show that the instantaneous accuracy is in the range of ±5-10% for values around 50%. It has not been studied how this, plus errors in the 3-hourly PERSIANN-CDR precipitation, affect the daily (precipitation-weighted) average, but we suspect that the average has better accuracy than the instantaneous estimates.

4.2.2 Error Budget

It is extremely challenging to develop an error budget for satellite retrievals. Error estimation is a current subject of active investigation and will be implemented in a future release.

The PLP field is being characterized. Early results tend to show that low(high) half-hourly probabilities are low(high). That is, the statements of phase are generally too confident. In addition, the lack of sub-daily TOVS/AIRS data introduces a difference in definition from that in the IMERG region. Finally for IMERG regions, inaccuracies in the joint time series of precipitation and phase could cause a somewhat inaccurate weighted phase estimate.

5.0 Practical Considerations

5.1 Quality Assessment and Diagnostics

Diagnostic analyses and plots of the output products are computed for selected months of days, with comparison to climatologies and the prior GPCP V1.3 1DD datasets. Particular attention is paid to the time series of various large-area averages for possible deviations as data boundaries are encountered. Anomalies are identified and analyzed to determine their origins.

5.2 Exception Handling

Errors in input data or processing are corrected when possible, or documented if they are not fixable.

5.3 Algorithm Validation

The diagnostics and exception handling described above form the first line of validation. Further evaluation will include validation against independent satellite-based precipitation data sets as well as in situ observations such as those from the Multi-Radar/Multi-Sensor System and Pacific atoll rain gauges.

5.4 Processing Environment and Resources

The computer used to process the GPCP Daily product is a CentOS 7 Linux Server. The programming languages and software include: C shell scripts to run the processing code, FORTRAN programs to perform the calculations, and xmgrace and Python to make the diagnostic plots and visualizations.

6.0 Assumptions and Limitations

There are a number of known issues that are relevant for a CDR-like data set. The GPCP team has worked hard to ameliorate these issues:

a. The initial release, labeled V3.2, has known limitations. Specifically, the TOVS/AIRS record is not as homogeneous as we expect for a CDR. The team continues to work toward improving these issues in a future release.
b. IMERG estimates are currently used only in the band 55°N-S, with the more-approximate TOVS/AIRS filling the polar regions. Further work is needed to explore the treatment of the actual spatially and temporally varying data boundary between IMERG and TOVS/AIRS.
c. Unlike the previous Version 1 and 2 GPCP SG datasets, the IR Tb data for GPCP V3.2 Monthly are provided in a consistent data format for the entire record.
d. TOVS data were partially denied for the period 10-18 September 2001 and cannot be recovered. As well, various operational issues caused partially or completely missing days of TOVS data.
e. Beginning with September 2002, AIRS precipitation estimates replaced the TOVS estimates at high latitudes because of TOVS product degradation later in 2003.
f. For December 1999-August 2002 (starting in June 2000), TOVS daily are calibrated to the zonal average AIRS-IR daily using two separate independent ENSO-neutral 24-month periods. However, regional differences remain.
g. Every attempt has been made to create an observation-only based precipitation data set. However, the TOVS estimates (but not AIRS) rely on numerical model data to initialize the estimation technique. The greatest chance of model influence is at high latitudes, where the retrievals more often fail to converge, and so fall back on the first guess. As well, the precipitation phase variable is a diagnostic based strictly on the MERRA-2 global analysis. This is believed to have only modest impact from numerical effects, since temperature and humidity are typically well-constrained by observations in MERRA-2.
h. Some polar-orbiting satellites have experienced significant drifting of the equator-crossing time during their period of service. There is no direct effect on the accuracy of the retrievals, but it is possible that the systematic change in sampling time could introduce biases in the resulting precipitation estimates. Satellites carrying the TOVS sensors did drift,

and a diurnal correction was applied to the data by the SRT. The Aqua satellite carrying the AIRS sensor has been station-keeping at 1:30 p.m. as part of the A-Train.

i. The PERSIANN-CDR contains artifacts in the Indian and Atlantic Oceans, apparently due to intersatellite calibration issues in the underlying GridSat IR Tb data. As such, there might slight leakage of this problem from the GPCP V3.2 Monthly to the Daily.

j. There are days when the TOVS or AIRS data are unavailable, and in such cases the areas outside the IR-based domain (55°N-S) are set to missing. One example is the span 1-18 November 2003.

7.0 Future Enhancements

The immediate plan is to compute additional months episodically when a critical mass of input data become available, which includes the shift to Version 07 IMERG. We will continue addressing the known anomalies in V3.2, namely inhomogeneities in the TOVS/AIRS record. In addition, there is a plan to extend the Daily data set for the entire 1983 – near-present period on the same 0.5° grid as for the Monthly. The fields will be globally complete for the period when daily TOVS/AIRS data are available (October 1996 – near-present), and over the latitude band 60°N-S for earlier times. As well, a 3-hourly product is anticipated, likely to be created by rescaling the IMERG half-hourlies (Huffman et al. 2020). This short-interval product is likely to be computed at the same 0.5° grid over the global domain for June 2000 – near-present (and shifting to January 1998 when all necessary inputs become available). Refining the methods for uncertainty quantification is also among the future tasks. Finally, the MEaSUREs-2017 project led by Dr. A. Behrangi (University of Arizona), within which V3.2 is being produced, is working toward improved high-latitude estimates. This can include implementation of the new retrievals and sensors as they become available and revisiting the gauge undercatch correction factors.

8.0 References

Behrangi, A., M. Lebsock, S. Wong, B. Lambrigtsen, 2012: On the Quantification of Oceanic Rainfall using Spaceborne Sensors. *J. Geophys.l Res.: Atmos.*, **117**. *doi:10.1029/2012jd017979*

Behrangi, A., G. Stephens, R.F. Adler, G.J. Huffman, B. Lambrigtsen, M. Lebsock, 2014: An Update on the Oceanic Precipitation Rate and Its Zonal Distribution in Light of Advanced Observations from Space. *J. Climate,* **27**, 3957-3965. *doi:10.1175/jcli-d-13-00679.1*

Behrangi, A., Y. Song, 2020: A New Estimate for Oceanic Precipitation Amount and Distribution Using Complementary Precipitation Observations from Space and Comparison with GPCP. *Environ. Res. Lett.*, **15**. *doi:10.1088/1748-9326/abc6d1*

Gelaro, R., W. McCarty, M.J. Suárez, R. Todling, A. Molod, L. Takacs, C.A. Randles, A. Darmenov, M.G. Bosilovich, R. Reichle, K. Wargan, L. Coy, R. Cullather, C. Draper, S. Akella. V. Buchard, A. Conaty, A.M. da Silva, W. Gu, G.-K. Kim, R. Koster, R. Lucchesi, D.Merkova, J.E. Nielsen, G. Partyka, S. Pawson, W. Putman, M. Rienecker, S.D. Schubert, M. Sienkiewicz, B. Zhao, 2017: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Climate*, **30,** 5419–5454. *doi:10.1175/JCLI-D-16-0758.1*

Hersbach, H, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, C. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, M. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R.J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, J.-N. Thépaut, 2020: The ERA5 Global Reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999-2049. doi:10.1002/qj.3803

Hong, Y., K.-L. Hsu, S. Sorooshian, X. Gao, 2004: Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System. *J. Appl. Meteor.*, **43**, 1834–1852. *doi:10.1175/JAM2173.1*

Huffman, G.J., R.F. Adler, A. Behrangi, D.T. Bolvin, E.J. Nelkin, M.R. Ehsani, 2022: Algorithm Theoretical Basis Document (ATBD) for Global Precipitation Climatology Project Version 3.2 Monthly Precipitation Data. MEaSUREs project, Greenbelt, MD, 31 pp. Available online: *https://docserver.gesdisc.eosdis.nasa.gov/public/project/MEaSUREs/GPCP/GPCP_ATBD_V3.2_ Monthly.pdf* (last accessed: January 21, 2023).

Huffman, G.J., R.F. Adler, M. Morrissey, D.T. Bolvin, S. Curtis, R. Joyce, B McGavock, J. Susskind, 2001: Global Precipitation at One-Degree Daily Resolution from Multi-Satellite Observations. *J. Hydrometeor.,* **2**, 36-50. *doi:10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2*

Huffman, G.J., D.T. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, C. Kidd, E.J. Nelkin, S. Sorooshian, E.F. Stocker, J. Tan, D.B. Wolff, P. Xie, 2020: Integrated Multi-satellitE Retrievals for the Global Precipitation Measurement (GPM) mission (IMERG). Chapter 19 in Adv. Global Change Res., Vol. 67, *Satellite Precipitation Measurement*, V. Levizzani, C. Kidd, D. Kirschbaum, C. Kummerow, K. Nakamura, F.J. Turk (Ed.), Springer Nature, Dordrecht, ISBN 978-3-030-24567-2 / 978-3-030-24568-9 (eBook), 343-353. *doi:10.1007/978-3-030-24568-9_19*

Joyce, R.J., P. Xie, J.E. Janowiak, 2011: Kalman Filter Based CMORPH. *J. Hydrometeor.*, **12**, 1547–1563. *doi:10.1175/JHM-D-11-022.1*

Kummerow, C.D., D.L. Randel, M. Kulie, N.-Y. Wang, R. Ferraro, S.J. Munchak, V. Petkovic, 2015: The Evolution of the Goddard PROFiling Algorithm to a Fully Parametric Scheme. *J. Atmos. Oc. Tech.*, **32**, 2265–2280. *doi:10.1175/JTECH-D-15-0039.1*

Olson, W.S., H. Masunaga, the GPM Combined Radar-Radiometer Algorithm Team, 2011: GPM Combined Radar-Radiometer Precipitation. Algorithm Theoretical Basis Document (Version 2). PPS, NASA/GSFC, 58 pp. Available at https://pps.gsfc.nasa.gov/Documents/GPM2011CombinedL2ATBD.pdf, last accessed 21 January 2023.

Sims, E.M., and G. Liu, 2015: A Parameterization of the Probability of Snow–Rain Transition. *J. Hydrometeor.*, **16**, 1466–1477. *doi:10.1175/JHM-D-14-0211.1*

Susskind, J. C.D. Barnet, J.M. Blaisdell, 2003: Retrieval of Atmospheric and Surface Parameters from AIRS/AMSU/HSB Data in the Presence of Clouds. *IEEE Trans. Geosci. Rem. Sens.*, **41**, 390-409. *doi:10.1109/TGRS.2002.808236*

Susskind, J., J. Pfaendtner, 1989: Impact of Interactive Physical Retrievals on NWP. *Report on the Joint ECMWF/EUMETSAT Workshop on the Use of Satellite Data in Operational Weather Prediction: 1989–1993*, *Vol. 1*, T. Hollingsworth, Ed., ECMWF, Shinfield Park, Reading RG2 9AV, U.K., 245-270.

Susskind, J., P. Piraino, L. Rokke, L. Iredell, A. Mehta, 1997: Characteristics of the TOVS Pathfinder Path A Dataset. *Bull. Amer. Meteor. Soc.*, **78**, 1449-1472.

Wang, X., Y. Feng, E. Mekis, 2017: Adjusted Daily Rainfall and Snowfall Data for Canada. *Atmos.-Ocean*, **55**, 155-168. *doi:10.1080/07055900.2017.1342163*

Appendix A. Acronyms and Abbreviations

1DD	One-Degree Daily
2BCMB	GPM Level 2 Combined radar-radiometer product
2BCMBT	TRMM Level 2 Combined radar-radiometer product
AIRS	Atmospheric Infrared Sounder
ATBD	Algorithm Theoretical Basis Document
CDR	Climate Data Record
CMORPH-KF	CPC Morphing - Kalman Filter
CORRA	Combined Radar-Radiometer Algorithm
CORRA-T	Combined Radar-Radiometer Algorithm – TRMM
CPC	Climate Prediction Center
DMSP	Defense Meteorological Satellite Program
DWD	Deutscher Wetterdienst
ENSO	El Niño Southern Oscillation
ERA5	European Centre for Medium-Range Weather Forecasting (ECMWF) Reanalysis 5
GES DISC	Goddard Earth Science Data and Information Services Center
GEWEX	Global Water and Energy Exchange project
GMI	GPM Microwave Imager
GPCC	Global Precipitation Climatology Centre
GPCP	Global Precipitation Climatology Project
GPM	Global Precipitation Measurement mission

GPROF	Goddard Profiling retrieval algorithm
GSFC	Goddard Space Flight Center
HIRS2	High-Resolution Infrared Sounder 2
IMERG	Integrated Multi-satellitE Retrievals for GPM
IR	infrared
MCTG	Merged CloudSat, Tropical Rainfall Measurement Mission (TRMM), and Global Precipitation Measurement (GPM) mission Climatology
MEaSUREs	Making Earth Science Data Records for Use in Research Environments
MERRA2	Modern-Era Retrospective analysis for Research and Applications Version 2
METH	Microwave Emission Brightness Temperature Histograms
MSU	Microwave Sounding Unit
NASA	National Aeronautics and Space Administration
NetCDF	Network Common Data Format
NOAA	National Oceanic and Atmospheric Administration
PERSIANN-CCS	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks – Cloud Classification System
PERSIANN-CDR	PERSIANN Climate Data Record
PLP	Probability of Liquid Phase
PMW	passive microwave
SG	Satellite-Gauge
SRT	Sensor Research Team
SSMI	Special Sensor Microwave/Imager
SSMIS	Special Sensor Microwave Imager/Sounder
SWE	snow water equivalent
TCC	TRMM Combined Climatology
TOVS	Television InfraRed Operational Satellite (TIROS) Operational Vertical Sounder
TRMM	Tropical Rainfall Measuring Mission
WCRP	World Climate Research Programme

Appendix B. Data Set Sources

<u>TOVS</u>
Legacy data currently only available upon request from George Huffman (george.j.huffman@nasa.gov)

<u>AIRS</u>
https://disc.gsfc.nasa.gov/datasets/AIRG2SSD_IRonly_006/summary

<u>IMERG Final</u>

https://disc.gsfc.nasa.gov/datasets/GPM_3IMERGHH_06/summary?keywords=imerg

GPCP V3.2 Monthly
https://doi.org/ 10.5067/MEASURES/GPCP/DATA304